Atty. Dkt. No. 071671/0156

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: Tadashi EMORI, et al.

Title: SPEAKER'S VOICE RECOGNITION SYSTEM, METHOD AND RECORDING MEDIUM

Appl. No.: Unassigned

Filing Date: October 25, 2000

Examiner: Unassigned

Art Unit: Unassigned

CLAIM FOR CONVENTION PRIORITY

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

The benefit of the filing date of the following prior foreign application filed in the following foreign country is hereby requested, and the right of priority provided in 35 U.S.C. § 119 is hereby claimed.

In support of this claim, filed herewith is a certified copy of said original foreign application:

Japanese Patent Application
No. 11-304685 filed 26 OCTOBER 1999.

Respectfully submitted,

Date: October 25, 2000

By Phillip J. Articola for / David A. Blumenthal
Attorney for Applicant
Registration No. 26,257
Reg. No. 38,819

FOLEY & LARDNER
Washington Harbour
3000 K Street, N.W., Suite 500
Washington, D.C. 20007-5109
Telephone: (202) 672-5407
Facsimile: (202) 672-5399

日 本 国 特 許 庁
PATENT OFFICE
JAPANESE GOVERNMENT

US

別紙添付の書類に記載されている事項は下記の出願書類に記載されてい
る事項と同一であることを証明する。
This is to certify that the annexed is a true copy of the following application as filed
th this Office.

出願年月日
Date of Application: １９９９年１０月２６日

出願番号
pplication Number: 平成１１年特許願第３０４６８５号

願人
plicant (s): 日本電気株式会社

CERTIFIED COPY OF
PRIORITY DOCUMENT

２０００年 ７月１４日

特許庁長官
Commissioner,
Patent Office 及川耕造

出証番号 出証特２０００－３０５４４６４

【書類名】　特許願

【整理番号】　33509635

【提出日】　平成11年10月26日

【あて先】　特許庁長官　殿

【国際特許分類】　G10L　3/00

G10L　3/02

【発明者】

【住所又は居所】　東京都港区芝五丁目７番１号　日本電気株式会社内

【氏名】　江森　正

【発明者】

【住所又は居所】　東京都港区芝五丁目７番１号　日本電気株式会社内

【氏名】　篠田　浩一

【特許出願人】

【識別番号】　000004237

【氏名又は名称】　日本電気株式会社

【代理人】

【識別番号】　100080816

【弁理士】

【氏名又は名称】　加藤　朝道

【電話番号】　045-476-1131

【手数料の表示】

【予納台帳番号】　030362

【納付金額】　21,000円

【提出物件の目録】

【物件名】　明細書　1

【物件名】　図面　1

【物件名】　要約書　1

【包括委任状番号】　9304371

【プルーフの要否】　要

【書類名】　明細書

【発明の名称】　音声認識装置及び方法ならびに記録媒体

【特許請求の範囲】

【請求項１】

音声信号のスペクトルを周波数軸上で伸縮を行うスペクトル変換部が、

入力音声信号をケプストラムを含む入力パターンに変換する分析部と、

標準パターンを記憶する標準パターン記憶部と、

前記入力パターンと前記標準パターンとを用いて周波数軸方向の伸縮パラメータを出力する伸縮推定部と、

前記伸縮パラメータを用いて前記入力パターンを変換する変換部と、

を備えたことを特徴とする音声認識装置。

【請求項２】

入力音声信号をケプストラムを含む入力パターンに変換する分析部と、

標準パターンを記憶する標準パターン記憶部と、

前記入力パターンと前記標準パターンとを用いて周波数軸方向の伸縮パラメータを出力する伸縮推定部と、

前記伸縮パラメータを用いて前記入力パターンを変換する変換部と、

前記変換部から出力された伸縮後の前記入力パターンと前記標準パターンとの距離を算出し最も距離の小さい前記標準パターンを認識結果として出力する整合部と、

を備えたことを特徴とする音声認識装置。

【請求項３】

前記変換部が、スペクトルの伸縮の形態を定めるワーピング関数によるスペクトルの周波数軸上の伸縮をケプストラム空間で行うことで、スペクトルの伸縮を行う、ことを特徴とする請求項１又は２記載の音声認識装置。

【請求項４】

前記伸縮推定部が、スペクトルの伸縮の形態を定めるワーピング関数によって定められたスペクトルの周波数軸上の伸縮を、ケプストラム空間におけるＨＭＭ（隠れマルコフモデル）の最尤推定から導出される推定を用いて伸縮パラメータ

の推定を行う、ことを特徴とする請求項１乃至３のいずれか一に記載の音声認識装置。

【請求項５】

学習用音声を記憶しておく学習音声記憶部と、

前記学習音声記憶部から学習用音声信号を入力し前記学習用音声信号をケプストラムを含む入力パターンに変換する分析部と、

標準パターンを記憶する標準パターン記憶部と、

前記入力パターンと前記標準パターンとを用いて周波数軸方向の伸縮パラメータを出力する伸縮推定部と、

前記伸縮パラメータを用いて入力パターンを変換する変換部と、

前記学習用音声に対し、前記変換部が出力した伸縮後の入力パターンと前記標準パターンとを用いて、前記標準パターン記憶部の標準パターンを更新する標準パターン推定部と、

前記変換部から出力される伸縮後の入力パターンと前記標準パターンとを用いて、距離を算出し、距離の変化量を監視する尤度判定部と、

を備えたことを特徴とする標準パターン学習装置。

【請求項６】

前記変換部が、スペクトルの伸縮の形態を定めるワーピング関数によるスペクトルの周波数軸上の伸縮をケプストラム空間で行うことでスペクトルの伸縮を行う、ことを特徴とする請求項５記載の標準パターン学習装置。

【請求項７】

前記伸縮推定部が、スペクトルの伸縮の形態を定めるワーピング関数によって定められたスペクトルの周波数軸上の伸縮をケプストラム空間でのＨＭＭの最尤推定から導出された推定方法を用いて伸縮パラメータの推定を行うことを特徴とする請求項５又は６記載の標準パターン学習装置。

【請求項８】

入力音声信号をケプストラムを含む入力パターンに変換する分析部と、

標準パターンを記憶する標準パターン記憶部と、

前記入力パターンと前記標準パターンとを用いて周波数軸方向の伸縮パラメー

タを出力する伸縮推定部と、

前記伸縮パラメータを用いて入力パターンを変換する変換部と、

前記変換部から出力された伸縮後の前記入力パターンの時系列を逆変換して時間領域波形を出力する逆変換部と、

を備えたことを特徴とする声質変換装置。

【請求項９】

音声信号のスペクトルを周波数軸上で伸縮を行うスペクトル変換部を構成するコンピュータにおいて、

（ａ）入力音声信号をケプストラムを含む入力パターンに変換する分析処理と、

（ｂ）前記入力パターンと、標準パターンを記憶する標準パターン記憶部に記憶されている前記標準パターンとを用いて周波数軸方向の伸縮パラメータを出力する伸縮推定処理と、

（ｃ）前記伸縮パラメータを用いて前記入力パターンを変換する変換処理と、

の前記（ａ）乃至（ｃ）の処理を前記コンピュータで実行させるためのプログラムを記録する記録媒体。

【請求項１０】

音声信号のスペクトルを周波数軸上で伸縮を行い音声認識を行う装置を構成するコンピュータにおいて、

（ａ）入力音声信号をケプストラムを含む入力パターンに変換する分析処理と、

（ｂ）前記入力パターンと、標準パターンを記憶する標準パターン記憶部に記憶されている前記標準パターンとを用いて周波数軸方向の伸縮パラメータを出力する伸縮推定処理と、

（ｃ）前記伸縮パラメータを用いて前記入力パターンを変換する変換処理と、

（ｄ）前記出力された伸縮後の前記入力パターンと前記標準パターンとの距離を算出し最も距離の小さい前記標準パターンを認識結果として出力する整合処理と、

の前記（ａ）乃至（ｄ）の処理を前記コンピュータで実行させるためのプログ

ラムを記録した記録媒体。

【請求項１１】

請求項１０記載の記録媒体において、前記変換処理が、スペクトルの伸縮の形態を定めるワーピング関数によるスペクトルの周波数軸上の伸縮をケプストラム空間で行うことで、スペクトルの伸縮を行う処理を前記コンピュータで実行させるためのプログラムを記録した記録媒体。

【請求項１２】

請求項１０記載の記録媒体において、前記伸縮推定処理が、スペクトルの伸縮の形態を定めるワーピング関数によって定められたスペクトルの周波数軸上の伸縮を、ケプストラム空間におけるＨＭＭ（隠れマルコフモデル）の最尤推定から導出される推定を用いて伸縮パラメータの推定を行う、処理を前記コンピュータで実行させるためのプログラムを記録した記録媒体。

【請求項１３】

学習用音声から標準パターンを学習する装置を構成するコンピュータにおいて、

（ａ）学習用音声を記憶しておく学習音声記憶部に記憶されている学習用音声を入力しケプストラムを含む入力パターンに変換する分析処理と、

（ｂ）前記入力パターンと、標準パターンを記憶する標準パターン記憶部に記憶されている前記標準パターンとを用いて周波数軸方向の伸縮パラメータを出力する伸縮推定処理と、

（ｃ）前記伸縮パラメータを用いて前記入力パターンを変換する変換処理と、

、

（ｄ）学習用音声に対し、変換処理が出力した伸縮後の入力パターンと、標準パターンを利用して、前記標準パターンを更新する標準パターン推定処理と、

（ｅ）前記変換処理で変換された伸縮後の入力パターンと前記標準パターンを利用して、距離を算出し、距離の変化量を監視する尤度判定処理と、

の前記（ａ）乃至（ｅ）の処理を前記コンピュータで実行させるためのプログラムを記録する記録媒体。

【請求項１４】

請求項１３記載の記録媒体において、前記変換処理が、スペクトルの伸縮の形態を定めるワーピング関数によるスペクトルの周波数軸上の伸縮をケプストラム空間で行うことで、スペクトルの伸縮を行う処理を前記コンピュータで実行させるためのプログラムを記録した記録媒体。

【請求項１５】

請求項１３記載の記録媒体において、前記伸縮推定処理が、スペクトルの伸縮の形態を定めるワーピング関数によって定められたスペクトルの周波数軸上の伸縮を、ケプストラム空間におけるＨＭＭ（隠れマルコフモデル）の最尤推定から導出される推定を用いて伸縮パラメータの推定を行う、処理を前記コンピュータで実行させるためのプログラムを記録した記録媒体。

【請求項１６】

音声信号のスペクトルを周波数軸上で伸縮を行う装置を構成するコンピュータにおいて、

（ａ）入力音声信号をケプストラムを含む入力パターンに変換する分析処理と、

（ｂ）前記入力パターンと、標準パターンを記憶する標準パターン記憶部に記憶されている前記標準パターンとを用いて周波数軸方向の伸縮パラメータを出力する伸縮推定処理と、

（ｃ）前記伸縮パラメータを用いて前記入力パターンを変換する変換処理と、

（ｄ）前記出力された伸縮後の前記入力パターンの時系列を逆変換して時間領域波形を出力する逆変換処理と、

の前記（ａ）乃至（ｄ）の処理を前記コンピュータで実行させるためのプログラムを記録する記録媒体。

【請求項１７】

音声信号のスペクトルを周波数軸上で伸縮を行うスペクトル変換を行う音声認識方法において、

入力音声信号をケプストラムを含む入力パターンに変換する第１のステップと、

前記入力パターンと、標準パターンを記憶する標準パターン記憶部に記憶され

ている前記標準パターンとを用いて周波数軸方向の伸縮パラメータを出力する第２のステップと、

前記伸縮パラメータを用いて前記入力パターンを変換する第３のステップと、

を含むことを特徴とするスペクトル変換方法。

【請求項１８】

入力音声信号をケプストラムを含む入力パターンに変換する第１のステップと、

前記入力パターンと、標準パターンを記憶する標準パターン記憶部に記憶されている前記標準パターンとを用いて周波数軸方向の伸縮パラメータを出力する第２のステップと、

前記伸縮パラメータを用いて前記入力パターンを変換する第３のステップと、

前記出力された伸縮後の前記入力パターンと前記標準パターンとの距離を算出し最も距離の小さい前記標準パターンを認識結果として出力する第４のステップと、

を含むことを特徴とする音声認識方法。

【請求項１９】

前記第３のステップにおいて、スペクトルの伸縮の形態を定めるワーピング関数によるスペクトルの周波数軸上の伸縮をケプストラム空間で行うことで、スペクトルの伸縮を行う、ことを特徴とする請求項１７又は１８記載の音声認識方法。

【請求項２０】

前記第２のステップにおいて、スペクトルの伸縮の形態を定めるワーピング関数によって定められたスペクトルの周波数軸上の伸縮を、ケプストラム空間におけるＨＭＭ（隠れマルコフモデル）の最尤推定から導出される推定を用いて伸縮パラメータの推定を行う、ことを特徴とする請求項１７乃至１９のいずれか一に記載の音声認識方法。

【請求項２１】

学習用音声を記憶しておく学習音声記憶部に記憶されている学習用音声を入力しケプストラムを含む入力パターンに変換する第１のステップと、

前記入力パターンと、標準パターンを記憶する標準パターン記憶部に記憶されている前記標準パターンとを用いて周波数軸方向の伸縮パラメータを出力する第２のステップと、

前記伸縮パラメータを用いて前記入力パターンを変換する第３のステップと、

前記学習用音声に対し、変換処理が出力した伸縮後の入力パターンと、標準パターンを利用して、前記標準パターンに更新する第４のステップと、

伸縮後の入力パターンと標準パターンを利用して、距離を算出し、距離の変化量を監視し尤度判定処理を行う第５のステップと、

を含むことを特徴とする標準パターン学習方法。

【請求項２２】

前記第３のステップにおいて、スペクトルの伸縮の形態を定めるワーピング関数によるスペクトルの周波数軸上の伸縮をケプストラム空間で行う、ことを特徴とする請求項２１記載の標準パターン学習方法。

【請求項２３】

前記第２のステップにおいて、スペクトルの伸縮の形態を定めるワーピング関数によって定められたスペクトルの周波数軸上の伸縮を、ケプストラム空間におけるＨＭＭ（隠れマルコフモデル）の最尤推定から導出される推定を用いて伸縮パラメータの推定を行う、ことを特徴とする請求項２１記載の標準パターン学習方法。

【請求項２４】

音声信号のスペクトルを周波数軸上で伸縮を行うスペクトル変換を行う音声認識方法において、

ワーピング関数で定義された、入力音声信号のスペクトルの伸縮を、ケプストラム上で行い、スペクトルの周波数軸の伸縮の度合いを、前記ワーピング関数に含まれる伸縮パラメータで決定し、伸縮パラメータの値を、話者毎に、最適な値を決定する、ことを特徴とする音声認識方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、不特定話者の音声認識装置、音声認識方法と音響モデルの学習方法および音声認識用プログラムを記録した記録媒体に関し、特に、周波数軸上の話者性を正規化できる音声認識装置および正規化学習装置、音声認識方法および正規化学習方法、音声認識用プログラムおよび正規化学習プログラムを記録した記録媒体に関する。

【０００２】

【従来の技術】

従来の音声認識装置のスペクトル変換部として、例えば特開平６－２１４５９６号公報（「文献１」という）や、１９９７年にアイ・イー・イー・イー（ＩＥＥＥ）から刊行された「インターナショナル・カンファレンス・オン・アコウスティックス・スピーチ・アンド・シグナル・プロシーディング」の論文集の第１０３９頁から第１０４２頁に掲載された「スピーカー・ノーマリゼーション・ベースド・オン・フリークエンシー・ワーピング」と題するザーン達による論文（Puming Zhan and Martin Westphal "Speaker NormalizationBased On Frequency Warping",ICASSP,1039-1042,1997）（「文献２」という）等の記載が参照される。

【０００３】

例えば、上記文献１には、予め定められた複数の異なる周波数特性補正係数に基づいて、入力された音声信号の周波数特性を補正する周波数補正手段と、予め定められた複数の周波数軸変換計数に基づいて、入力された音声信号の周波数軸を変換する周波数軸変換手段と、入力された音声信号の特徴量を入力音声特徴量として抽出する特徴量抽出手段と、標準音声特徴量を保持している標準音声記憶手段と、周波数特性補正手段、周波数軸変換手段、特徴量抽出手段により処理され得られた入力音声特徴量と標準音声記憶手段に保持されている標準音声特徴量とを照合する照合手段と、を備え、話者適応フェーズと音声認識フェーズの機能を有する音声認識装置であって、話者適応フェーズにおいて、既知なる発生内容の未知なる話者の入力音声信号に対して、複数の異なる周波数特性補正係数、複数の異なる周波数軸変換係数の各々の係数ごとに、周波数特性補正手段、周波数軸変換手段、特徴量抽出手段の処理を行わせて、各々の係数毎の入力音声特徴量

と既知なる発生内容と同一内容の標準音声特徴量と照合し、各々の係数のうちから、最小距離を与える１つの周波数特性補正係数と１つの周波数軸変換係数を選択し、音声認識フェーズでは、選択された周波数特性補正係数と周波数軸変換係数を用いて、入力音声特徴量を求め、これを標準音声特徴量と照合し音声認識処理を行う構成が提案されている。

【０００４】

これら従来の音声認識装置において、スペクトル変換部は、話者の性別や年齢や体格などの個人差に対して、音声信号のスペクトルを周波数軸上で伸縮を行うことにより、認識性能の改善を行う。

【０００５】

そしてスペクトル変換部では、周波数軸上で伸縮を行うために、適当なパラメータでその伸縮の概形を変えることのできる関数を定義し、音声信号のスペクトルを、ワーピング関数を用いて周波数軸上の伸縮を行う。

【０００６】

周波数軸上での伸縮を行うための関数を「ワーピング関数」という。また、ワーピング関数の概形を定義するパラメータを「伸縮パラメータ」という。

【０００７】

従来、ワーピング関数の伸縮パラメータ（「ワーピングパラメータ」と略記する）として、複数個のワーピングパラメータの値を用意し、１つ１つの値を用いて音声信号のスペクトルを周波数軸上で伸縮を行い、伸縮されたスペクトルを用いて入力パターンを計算し、入力パターンと標準パターンを用いて距離を求め、該距離が最小になる値を、認識時のワーピングパラメータの値としている。

【０００８】

従来の音声認識装置のスペクトル変換部について、図面を参照して以下に説明する。図９は、従来の音声認識装置のスペクトル変換部の構成の一例を説明するための図である。図９を参照すると、この従来のスペクトル変換部は、ＦＦＴ（Fast Fourier Transform；高速フーリエ変換）部３０１と、伸縮パラメータ記憶部３０２と、周波数変換部３０３と、入力パターン計算部３０４と、整合部３０６と、標準パターン３０５と、伸縮ラメータ選択部３０７と、を備えている。

【０００９】

ＦＦＴ部３０１は、入力された音声信号を一定の時間毎に切り出してフーリエ変換を行い、周波数スペクトルを求める。

【００１０】

伸縮パラメータ記憶部３０２は、周波数の伸縮を決定する伸縮パラメータの値を複数記憶している。

【００１１】

周波数変換部３０３は、伸縮パラメータによりその概形が決定されたワーピング関数を用い、ＦＦＴ部３０１から出力されるスペクトルに対して、周波数の伸縮を行い、周波数の伸縮が行われたスペクトルを伸縮スペクトルとして出力する。

【００１２】

入力パターン計算部３０４は、周波数変換部３０３から出力された伸縮スペクトルを用い、入力パターンを計算し、出力する。

【００１３】

入力パターンは、ケプストラムなど音響的特徴を表すパラメータの時系列などを表す。

【００１４】

標準パターンは、入力パターンなどを数多く用い、同じクラスに属する音素単位の入力パターンを、ある種の平均化の手法を用い作成されるものである。標準パターンの作成については、１９９５年にＮＴＴアドバンステクノロジ株式会社から刊行された「古井監訳、音声認識の基礎」（上）の第６３頁（「文献３」という）の記載が参照される。

【００１５】

標準パターンの種類には、認識アルゴリズムにより、例えばＤＰ（Dynamic Programming；ダイナミックプログラミング）マッチングの場合、音素の時系列順に入力パターン並べられた時系列標準パターン、HMM（Hidden Markov Model；隠れマルコフモデル）の場合、状態の系列やその接合情報となる。

【００１６】

整合部３０６は、ＦＦＴ部３０１に入力された発声の内容に合わせた標準パターン３０５と入力パターンを用い、距離を求める。算出される距離は、標準パターンがＨＭＭの場合、尤度に相当し、ＤＰマッチングの場合、最適経路の距離に相当する。

【００１７】

伸縮パラメータ選択部３０７は、整合部３０６で求められた整合性から最も整合性のとれている伸縮パラメータを選択する。

【００１８】

図１０は、従来のスペクトル整合部の処理を説明するための流れ図である。図０及び図１０を参照して、従来のスペクトル変換部の動作について説明する。

【００１９】

ＦＦＴ部３０１は、音声信号をＦＦＴ演算することにより、スペクトルを求め出力する（図１０のステップＤ１０１）。

【００２０】

周波数変換部３０３は、入力された伸縮パラメータ（Ｄ１０６）を用いてスペクトルの周波数軸の伸縮を行う（ステップＤ１０２）。

【００２１】

入力パターン計算部３０４は、周波数軸の伸縮されたスペクトルを用い、入力パターンを計算する（ステップＤ１０３）。

【００２２】

整合部３０５は、標準パターン（Ｄ１０７）と入力パターンの距離を求める（Ｄ１０４）。

【００２３】

ステップＤ１０１からＤ１０４までの一連の処理を、伸縮パラメータ記憶部３０２に記憶されている全ての伸縮パラメータの値について行う（ステップＤ１０５）。

【００２４】

すなわち、伸縮パラメータ記憶部３０２に、伸縮パラメータの値として１０個の値が記憶されている場合、ステップＤ１０１からＤ１０４の処理を１０回繰り

返し、１０通りの距離を求める。

【００２５】

伸縮パラメータ選択部３０７は、全ての伸縮パラメータに対応する距離を比較し、最も距離の小さい場合の、伸縮パラメータを選択する（ステップＤ１０８）。

【００２６】

【発明が解決しようとする課題】

しかしながら、上記した従来のスペクトル変換部は、下記記載の問題点を有している。

【００２７】

第１の問題点は、伸縮パラメータの値を決定する際に、演算量が大きくなる、ということである。

【００２８】

その理由は、従来のスペクトル変換部においては、伸縮パラメータの値を複数個用意しておき、その数の分だけ、

・ＦＦＴ、

・スペクトルの周波数の伸縮、

・入力パターンの計算と、

・距離を計算する処理

などを繰り返し行う必要があるためである。

【００２９】

第２の問題点は、周波数の伸縮による認識装置などへの効果が、十分得られない可能性が有る、ということである。

【００３０】

その理由は、伸縮パラメータの値は、予め定められた値しか持たないので、未知の話者に対して、それらの値の中に最適なものがあるとは限らないためである。

【００３１】

したがって、本発明は、上記問題点に鑑みてなされたものであって、その主た

る目的は、少ない演算量で話者毎に最適な伸縮パラメータの値を計算可能とし、性能を向上する音声認識装置及び方法並びに記録媒体を提供することにある。これ以外の本発明の目的、特徴等は以下の説明で直ちに明らかとされるであろう。

【００３２】

【課題を解決するための手段】

前記目的を達成する本発明は、入力音声信号をケプストラムを含む入力パターンに変換する分析部と、標準パターンを記憶する標準パターン記憶部と、前記入力パターンと前記標準パターンとを用いて周波数軸方向の伸縮パラメータを出力する伸縮推定部と、前記伸縮パラメータを用いて前記入力パターンを変換する変換部と、前記変換部が出力した伸縮後の前記入力パターンと前記標準パターンとの距離を算出し、最も距離の小さい前記標準パターンを認識結果として出力する整合部と、を備えている。

【００３３】

本発明は、学習用音声を記憶しておく学習音声記憶部と、前記学習音声記憶部から学習用音声信号を入力し前記学習用音声信号をケプストラムを含む入力パターンに変換する分析部と、標準パターンを記憶する標準パターン記憶部と、前記入力パターンと前記標準パターンを用いて周波数軸方向の伸縮パラメータを出力する伸縮推定部と、前記伸縮パラメータを用いて入力パターンを変換する変換部と、前記標準パターンを記憶する標準パターン記憶部と、学習用音声に対し、変換部が出力した伸縮後の入力パターンと標準パターンを利用して、標準パターンに更新する標準パターン推定部と、伸縮後の入力パターンと標準パターンを利用して、距離を算出し、距離の変化量を監視する尤度判定部と、を備える。

【００３４】

本発明は、入力音声信号をケプストラムを含む入力パターンに変換する第１の工程と、前記入力パターンと、標準パターンを記憶する標準パターン記憶部に記憶されている前記標準パターンとを用いて周波数軸方向の伸縮パラメータを出力する第２の工程と、前記伸縮パラメータを用いて前記入力パターンを変換する第３の工程と、前記出力された伸縮後の前記入力パターンと前記標準パターンとの距離を算出し最も距離の小さい前記標準パターンを認識結果として出力する第４

の工程と、を含む。

【００３５】

【発明の実施の形態】

本発明の実施の形態について説明する。本発明は、入力音声信号をケプストラムを含む入力パターンに変換する分析部（１）と、標準パターンを記憶する標準パターン記憶部（４）と、入力パターンと標準パターンを用いて周波数軸方向の伸縮パラメータを出力する伸縮推定部（３）と、前記伸縮パラメータを用いて入力パターンを変換する変換部（２）と、を備える。

【００３６】

また本発明は、変換部（２）によって変換された前記入力パターンと標準パターンとの距離を計算し、最も距離の小さい標準パターンを認識結果として出力する整合部（認識部１０１）を備える。

【００３７】

伸縮推定部（３）は、入力パターンに含まれるケプストラムを用いて、伸縮パラメータを推定する。このため、本発明によれば、伸縮パラメータを決定する際に、前もって様々な値を記憶保持しておく必要も無く、また様々な値について距離計算を行う必要もない。

【００３８】

また本発明の装置は、学習用音声を記憶しておく学習音声記憶部（２０１）と、学習音声記憶部（２０１）から学習用音声を入力しケプストラムを含む入力パターンに変換する分析部（１）と、標準パターンを記憶する標準パターン記憶部（４）と、前記入力パターンと前記標準パターンを用いて周波数軸方向の伸縮パラメータを出力する伸縮推定部（３）と、前記伸縮パラメータを用いて入力パターンを変換する変換部（２）と、前記標準パターンを記憶する標準パターン記憶部と、学習用音声に対し、変換部が出力した伸縮後の入力パターンと標準パターンを利用して、標準パターンに更新する標準パターン推定部（２０２）と、伸縮後の入力パターンと標準パターンを利用して、距離を算出し、距離の変化量を監視する尤度判定部（２０３）と、を備える。

【００３９】

【実施例】

次に、本発明の実施例について図面を参照して詳細に説明する。図１は、本発明の第１の実施例の音声認識装置のスペクトル変換部の構成を示す図である。図１を参照すると、本発明の第１の実施例をなす音声認識装置のスペクトル変換部は、分析部１と、変換部２と、伸縮推定部３と、標準パターン記憶部４とを備えて構成されている。

【００４０】

分析部１は、一定の時間毎に音声信号を切り出し、ＦＦＴ（高速フーリエ変換）やＬＰＣ（linear predictive coding;線形予測符号）分析などを用い、スペクトル成分を求め、人間の聴覚を考慮したメルスケールに変換し、そのメルスペクトルの成分の包絡成分を抽出するためのメルケプストラムを求め、そのメルケプストラムや、その変化量、変化量の変化量などを、入力パターンとして出力する。

【００４１】

変換部２は、周波数の伸縮を、入力パターン中のメルケプストラムを変換することにより行う。ここで、変換部２で行われる変換の一例について詳細に説明する。

【００４２】

次式（１）で示される１次の全域通過型のフィルタによる周波数の変換は、例えば、１９７２年にアイ・イー・イー・イー(IEEE)から発行された「プロシーディングス・オブ・アイ・イー・イー・イー」のボリューム６０の第６号の第６８１頁から６１９頁に掲載された「ディスクリート・リプレゼンテーション・オブ・シグナルス」と題するオッペンハイムによる論文(Oppenheim：”Discrete Representation of Signals,” Proc.IEEE,60,681-691,June 1972)（「文献４」という）によると、ケプストラム（記号c、添え数字はケプストラムの次元数とする）を用いた次式（２）で示される再帰的表現で表すことができる。

【００４３】

$$\hat{z}^{-1}=\frac{z^{-1}-\alpha}{1-\alpha z^{-1}} \qquad \cdots(1)$$

$$\hat{c}_m^{(i)}=\begin{cases} c_{-i}+\alpha\hat{c}_0^{(i-1)}, & m=0 \\ (1-\alpha^2)\hat{c}_{m-1}^{(i-1)}+\alpha\hat{c}_1^{(i-1)}, & m=1 \\ \hat{c}_{m-1}^{(i-1)}+\alpha(\hat{c}_m^{(i-1)}-\hat{c}_{m-1}^{(i)}), & m\geq 2 \end{cases}$$

$$i=-\infty,\cdots,-1,0. \qquad \cdots(2)$$

【００４４】

式（２）よるケプストラム空間での変換が、上式（１）によるスペクトルの周波数の変換と等価になる。

【００４５】

そのため、変換部１０２は、周波数の伸縮のために直接スペクトルを用いるのではなく、式（１）をワーピング関数とし、上式（１）中のαを伸縮パラメータとし、式（１）から導かれる式（２）による変換を、入力パターンに対して施すことにより、スペクトルの周波数の伸縮を行う。

【００４６】

変換された入力パターンは、変換入力パターンとして出力される。

【００４７】

標準パターン記憶部４には、標準パターンが記憶される。標準パターンは、単語や音素などを単位とした音韻情報として、隠れマルコフモデル（「HMM」という）や、音素の時系列などの時系列標準パターンなどでも代用できる。本実施例では、標準パターンは、隠れマルコフモデルとする。HMMを構成する情報は、連続ガウス分布における平均ベクトルや分散、さらに状態と状態の遷移確率等のなどがある。

【００４８】

伸縮推定部（「伸縮パラメータ推定部」ともいう）３は、分析部１に入力された音声信号に対応するHMMを用い、入力パターンとのアライメントを求める。ここで、アライメントとは、HMMの各時刻と各状態における事後確率のことである。

【００４９】

アライメントの求め方としては、例えば、１９９５年にＮＴＴアドバンステクノロジ株式会社から出版された「古井監訳、音声認識の基礎（下）」の第１０２頁乃至第１８５頁（「文献５」という）に記載されている、ビタビアルゴリズムやフォワードバックワードアルゴリズム等の公知の手法を用いて求めることができる。

【００５０】

求められたアライメントとHMMと入力パターンを用い、伸縮パラメータを計算する。伸縮パラメータの計算は、式（４）によって求められる。

【００５１】

$$\hat{c}_0 = \sum_{m=0}^{\infty} \alpha^m c_m,$$

$$\hat{c}_1 = (1-\alpha^2)\sum_{m=1}^{\infty} m\alpha^{m-1} c_m,$$

$$\hat{c}_2 = c_2 + \alpha(-c_1 + 3c_3) + \alpha^2(-4c_2 + 6c_4) + \cdots \alpha^3(c_1 - 9c_3 + 10c_5) + \cdots,$$

$$\hat{c}_3 = c_3 + \alpha(-2c_2 + 4c_4) + \alpha^2(c_1 - 9c_3 + 10c_5) + \alpha^3(6c_2 - 24c_4 + 20c_6) + \cdots, \quad \cdots(3)$$

$$\hat{c}_m = c_m + \begin{cases} (m+1)c_{m+1}\alpha, & m = 0 \\ \{(m+1)c_{m+1} - (m-1)c_{m-1}\}\alpha, & m > 0 \end{cases} \quad \cdots(4)$$

【００５２】

式（４）は、式（２）で表される再帰式を、式（３）のように、伸縮パラメータαについて展開し、αの１次の項で近似し、上記文献４に記載されているように、HMMの最尤推定のためのＱ関数に組み入れ、αについて、Ｑ関数を最大にすることで導くことができる。

【００５３】

導かれた関数が、式（５）である。

【００５４】

$$\alpha = \frac{\sum_{t=1}^{T}\gamma_t(j,k)\left[\sum_{m=1}^{M}\frac{1}{\sigma_m^2}\Delta c_{mt}\bar{c}_{mt}\right]}{\sum_{t=1}^{T}\gamma_t(j,k)\left[\sum_{m=1}^{M}\frac{1}{\sigma_m^2}\bar{c}_{mt}^2\right]}$$

$$\Delta c_{mt} = c_{mt} - \mu_{jkm}, \qquad \cdots(5)$$

$$\bar{c}_{mt} = (m-1)c_{(m-1)t} - (m+1)c_{(m+1)t}$$

【００５５】

式（５）において、ｃは、前記の入力パターンのうちのメルケプストラム部分を表し、μは、ＨＭＭの平均ベクトル表し、σは、ＨＭＭの分散を表し、γは、アライメント情報である時刻ｔ、状態ｊ、混合状態ｋにおける事後確率を表す。

【００５６】

事後確率は、フォワードバックワードアルゴリズムの場合、ある時刻状態における存在確率であり、ビタビアルゴリズムの場合、ある時刻状態に最適経路に存在している場合１、存在していない場合０となる。

【００５７】

本実施例では、ワーピング関数として式（１）をあげたが、本発明において、式（１）に限定されるものでなく、任意の式が適用可能である。また、式（５）を導くために、式（２）の１次近似を用いたが、２次以上の多次の近似を用いてもよい。

【００５８】

図２は、本発明の第１の実施例の処理を説明するための流れ図である。図１及び図２を参照して、本発明の第１の実施例の全体の動作について詳細に説明する。まず、音声信号と入力し（図２のステップＡ１０１）、分析部１で、入力パターンの計算を行う（ステップＡ１０２）。

【００５９】

伸縮推定部３が、分析部１から出力される入力パターンと、入力されたＨＭＭ（Ａ１０５）とを用いて、伸縮パラメータを計算する（ステップＡ１０３）。

【００６０】

分析部１から出力された入力パターンを、変換部２が、式（２）、式（３）、

式（４）のいずれかの変換の関数を用い、変換入力パターンを求める（ステップＡ１０４）。ただし、αの値は、初めての発声の場合０とし、２回目以降の発声の場合、伸縮推定部３が出力した値を用いる。

【００６１】

次に本発明の第１の実施例の作用効果について説明する。本発明の第１の実施例では、分析部１の出力である入力パターンが、変換部２に入力されるように構成されているため、スペクトルの周波数の伸縮を、メルケプストラム領域で行うことができる。

【００６２】

また、式（５）を用いた場合、従来の技術で説明したような、繰り返しの計算が無く、分析処理などが１回で済むため、伸縮パラメータを推定する際の演算量を少なくできる。

【００６３】

次に、本発明の第２の実施例について説明する。図３は、本発明の第２の実施例の構成を示す図である。図３を参照すると、本発明の第２の実施例をなす音声認識装置は、分析部１と、変換部２と、伸縮推定部３と、認識部１０１と、標準パターン記憶部４と、を備えている。分析部１、変換部２、伸縮推定部３、標準パターン記憶部４は、前記第１の実施例で説明したものと同様とされる。

【００６４】

すなわち、分析部１は、前記第１の実施例と同様、音声信号を分析し、入力パターンを計算し、出力する。

【００６５】

変換部２は、前記第１の実施例と同様に、入力パターンの変換を行い、変換入力パターンとして出力する。

【００６６】

標準パターン記憶部４には、前記第１の実施例と同様、音韻を表す要素として、入力パターンの平均ベクトルや分散等で構成されるHMMが記憶される。

【００６７】

認識部（整合部）１０１は、HMMのうち、いずれが、変換部から出力された

変換入力パターンによく整合するかを調べることにより、認識を行い、認識結果を出力する。整合の方法として、上記文献４で示されるビタビアルゴリズムやフォワードバックワードアルゴリズムのような公知の方法が用いられる。

【００６８】

図４は、本発明の第２の実施例の処理手順を説明するための流れ図である。図３及び図４を参照して、本発明の第２の実施例の全体の動作について詳細に説明する。

【００６９】

分析部１が、入力された音声信号（図４のステップＢ１０１）を分析し、入力パターンの計算を行う（ステップＢ１０２）。分析部１から出力された入力パターンを、変換部２が、式（２）、式（３）、式（４）のいずれかの変換の関数を用い、変換入力パターンを求める（ステップＢ１０３）。ただし、αの値は、初めての発声の場合０とし、２回目以降の発声の場合、ワーピングパラメータの伸縮推定部３が出力した値を用いる。

【００７０】

つぎに、変換入力パターンを用い、認識部１０１が、認識処理を行う（ステップＢ１０４）。

【００７１】

このとき、認識部１０１に、標準パターン記憶部４からHMMが入力される（ステップＢ１０６）。認識処理後、認識結果を基に伸縮パラメータの推定部００３が、伸縮パラメータを計算する（ステップＢ１０５）。

【００７２】

以後、ステップＢ１０５で求めた伸縮パラメータを用い、ステップＢ１０１の音声の入力の処理から繰り返す。

【００７３】

次に、本発明の第２の実施例の作用効果について説明する。本発明の第２の実施例では、前記第１の実施例のスペクトル変換部１００と認識部１０１とを備えて構成されており、音声信号が入力されるたびに、伸縮パラメータの値が更新され、標準パターンとの周波数のずれを補正することができるため、認識性能を向

上している。

【００７４】

また本発明の第２の実施例では、伸縮パラメータの推定をＨＭＭの最尤推定のＱ関数を最小にするような式（５）を用いて行っているため、伸縮パラメータが連続的な値で推定され、予め用意された離散的な値を用いた場合に比べ、認識性能の向上を期待することができる。

【００７５】

次に、本発明の第３の実施例について説明する。図５は、本発明の第３の実施例の構成を示す図である。図５を参照すると、本発明の第３の実施例をなす標準パターン学習装置は、前記第１の実施例のスペクトル変換部１００に加えて、学習音声記憶部２０１と、標準パターン推定部２０２と、尤度判定部２０３とを備えた構成されている。

【００７６】

学習用音声記憶部２０１は、ＨＭＭの学習に用いる音声信号を記憶する。

【００７７】

標準パターン推定部２０２は、スペクトル変換部１００が出力する変換入力パターンとＨＭＭを用い、ＨＭＭのパラメータを推定する。推定方法は、例えば、文献４等に記載されている、最尤推定を用いる。

【００７８】

尤度判定部２０３は、スペクトル変換部１００が出力する変換入力パターンとＨＭＭを用い、全ての学習用の音声信号に対し、距離を求める。距離を求める方法として、標準パターンがＨＭＭなら、上記文献５に記載されているビタビアルゴリズムやフォワードバックワードアルゴリズムのような方法が用いられる。

【００７９】

本発明の第３の実施例では、ＨＭＭの学習を例として説明しているが、音声認識関係のパラメータの学習であれば、どのようなものでもよい。

【００８０】

図６は、本発明の第３の実施例の処理動作を説明するための流れ図である。図５及び図６を参照して、本発明の第３の実施例の全体の動作について詳細に説明

する。まず、学習用の音声信号を、スペクトル変換部１００のスペクトル分析部１に入力する（図６のＣ１０２）。

【００８１】

分析部１は、学習用の信号を分析し入力パターンを出力する（ステップＣ１０２）。伸縮推定部００３が、伸縮パラメータを推定する（ステップＣ１０３）。

【００８２】

変換部２で、入力パターンの変換を行い、変換入力パターンを出力する（ステップＣ１０４）。

【００８３】

標準パターン推定部２０２において、変換入力パターンとＨＭＭを用い、ＨＭＭの推定を行う（ステップＣ１０５）。

【００８４】

尤度判定部２０３が、全ての音声信号に対する尤度を求め、尤度の変化量とその閾値を比較し、尤度の変化量が閾値以下の場合、標準パターン推定部２０２で推定されたＨＭＭで標準パターン記憶部４を更新して、学習を終了する。

【００８５】

一方、尤度の変化量が、閾値より大きい場合、標準パターン推定部２０２で推定されたＨＭＭによって標準パターン記憶部４を更新し、学習音声データの入力処理（Ｃ１０１）から、再び一連の処理を行う。

【００８６】

本発明の第３の実施例の作用効果について説明する。本発明の第３の実施例によれば、話者ごとの周波数の伸縮の影響を、ワーピング関数で修正した標準パターンを学習する際、伸縮パラメータの推定が学習処理の中でできるため、従来の技術よりも演算量を少なくできる。また、伸縮パラメータの推定に用いる式（５）は、ＨＭＭの最尤推定を用いて導出された式であるので、他のＨＭＭのパラメータの推定と同様に、学習の過程に組み入れることが容易である。

【００８７】

次に本発明の第４の実施例について説明する。図７は、本発明の第４の実施例の構成を示す図である。図７を参照すると、本発明の第４の実施例は、前記第１

の実施例の構成に加えて逆変換部５を備えている。逆変換部５は、変換部２から出力された伸縮後の前記入力パターンの時系列を逆変換して時間領域（time domain）の信号波形を出力することで、声質の変換を行う。

【００８８】

次に本発明の第５の実施例について説明する。図８は、本発明の第５の実施例の構成を示す図である。本発明の第５の実施例は、前記した第１乃至第４の実施例の装置を、コンピュータで実行されるプログラム制御によって実現するものである。図８を参照すると、図１に示した、分析部１、変換部２、伸縮推定部３の処理を、コンピュータ１０上でプログラムを実行することで実現する場合、プログラムを記録したＣＤ－ＲＯＭ、ＤＶＤ、ＦＤ、磁気テープ等の記録媒体１４から記録媒体アクセス装置１３を介してコンピュータ１０の主記憶１２にロードし、ＣＰＵ１１が該プログラムを実行する。すなわち、記録媒体１４には、入力音声信号をケプストラムを含む入力パターンに変換する分析処理と、前記入力パターンと、標準パターンを記憶する標準パターン記憶部に記憶されている前記標準パターンとを用いて周波数軸方向の伸縮パラメータを出力する伸縮推定処理と、前記伸縮パラメータを用いて前記入力パターンを変換する変換処理と、の各処理を前記コンピュータで実行させるためのプログラムが記録される。

【００８９】

また、前記出力された伸縮後の前記入力パターンと前記標準パターンとの距離を算出し最も距離の小さい前記標準パターンを認識結果として出力する整合処理をコンピュータで実行させるためのプログラムを記録してもよい。

【００９０】

さらに、記録媒体１４には、学習用音声を記憶しておく学習音声記憶部に記憶されている学習用音声を入力しケプストラムを含む入力パターンに変換する分析処理と、前記入力パターンと、標準パターンを記憶する標準パターン記憶部に記憶されている前記標準パターンとを用いて周波数軸方向の伸縮パラメータを出力する伸縮推定処理と、前記伸縮パラメータを用いて前記入力パターンを変換する変換処理と、学習用音声に対し、変換処理が出力した伸縮後の入力パターンと、標準パターンを利用して、前記標準パターンを更新する標準パターン推定処理と

、伸縮後の入力パターンと標準パターンを利用して、距離を算出し、距離の変化量を監視する尤度判定処理と、の各処理を前記コンピュータで実行させるためのプログラムを記録してもよい。このように、前記第２乃至第４の実施例についても、同様にして、プログラム制御で実現可能である。なお、プログラムは、不図示のサーバ装置からネットワーク等伝送媒体を介してダウンロードするようにしてもよい。すなわち、上記プログラムを担持する媒体であれば、記録媒体以外にも、通信媒体等任意の媒体を含む。

【００９１】

【発明の効果】

以上説明したように、本発明によれば下記記載の効果を奏する。

【００９２】

本発明の第１の効果は、音声信号のスペクトルの周波数の伸縮において、認識性能に最適なパラメータの計算に要する演算量を縮減する、ということである。

【００９３】

その理由は、本発明においては、周波数軸に対し１次の全域通過型等のフィルタ処理における変換が、ケプストラム領域で伸縮パラメータべき級数の形で解け、その級数を１次関数と近似したとき、最尤推定のためのＱ関数を最小にする伸縮パラメータの関数を容易な関数で記述することができ、該関数を用いて計算を行う構成としたためである。

【００９４】

本発明の第２の効果は、ＨＭＭの学習時に、他のパラメータと同時に伸縮パラメータを推定できる、ということである。

【００９５】

その理由は、本発明においては、前記第１の効果の理由で説明したように、伸縮パラメータを計算するための関数が、音声認識における最尤推定のためのＱ関数から導出されるためである。

【図面の簡単な説明】

【図１】

本発明の第１の実施例の構成を示す図である。

【図２】

本発明の第１の実施例の処理動作を説明するための流れ図である。

【図３】

本発明の第２の実施例の構成を示す図である。

【図４】

本発明の第２の実施例の処理動作を説明するための流れ図である。

【図５】

本発明の第３の実施例の構成を示す図である。

【図６】

本発明の第３の実施例の処理動作を説明するための流れ図である。

【図７】

本発明の第４の実施例の構成を示す図である。

【図８】

本発明の第５の実施例の構成を示す図である。

【図９】

従来のスペクトル変換部の構成を示す図である。

【図１０】

従来のスペクトル変換部の動作を示す流れ図である。

【符号の説明】

１　分析部

２　変換部

３　伸縮推定部

４　標準パターン記憶部

１００　スペクトル変換部

１０１　認識部

２０１　学習音声記憶部

２０２　標準パターン推定部

２０３　尤度判定部

３０１　ＦＦＴ部

３０２　伸縮パラメータ記憶部

３０３　周波数変換部

３０４　特徴パラメータ計算部

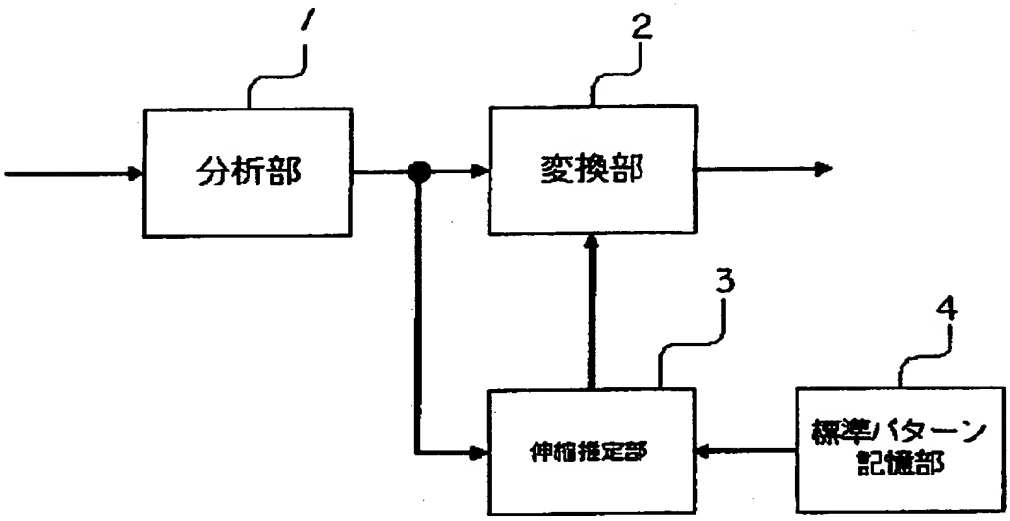
３０５　標準パターン

３０６　整合部

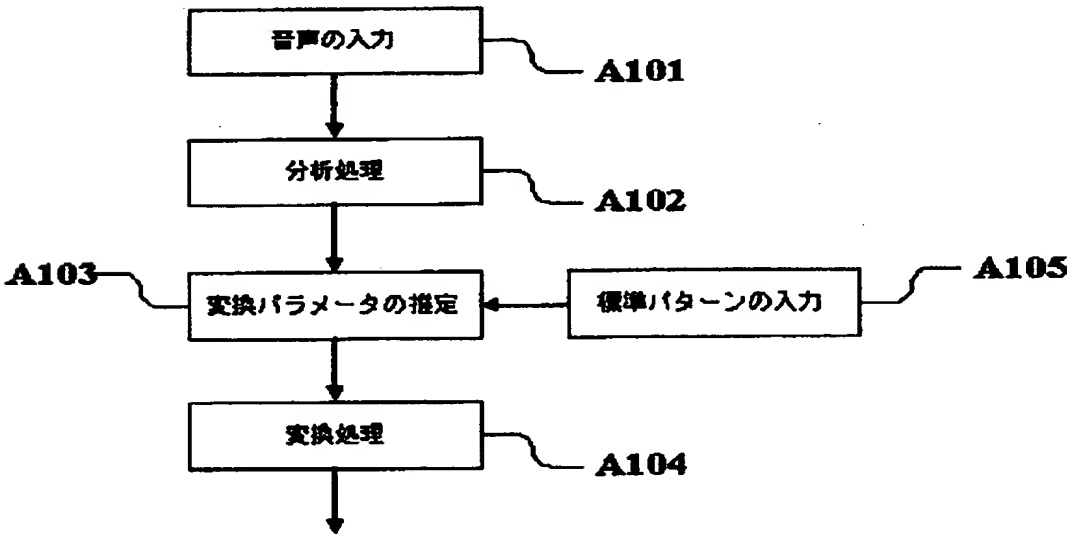
３０７　伸縮パラメータ選択部

【書類名】　　図面

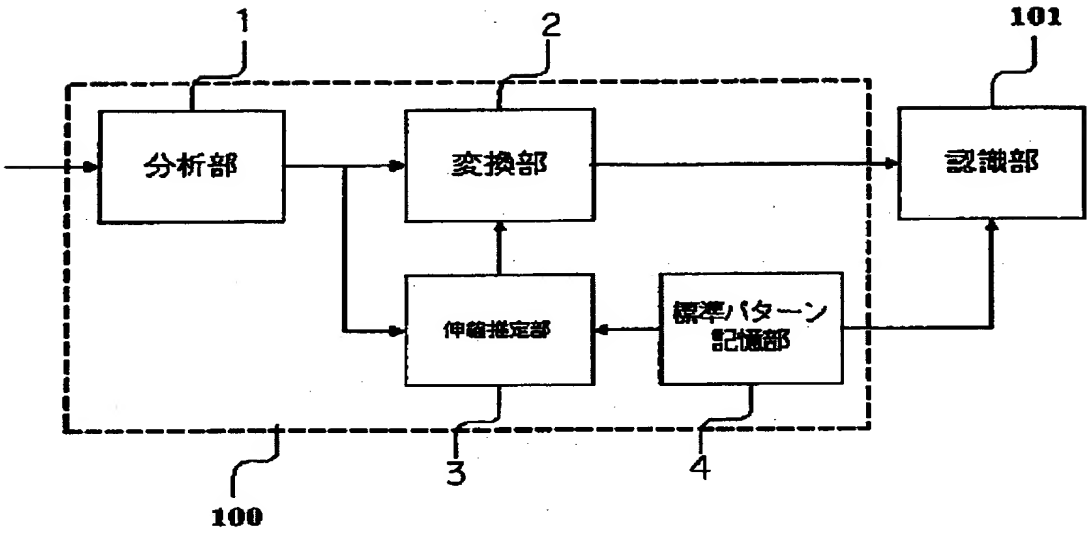
【図１】


1
2
分析部
変換部
3
4
伸縮推定部
標準パターン
記憶部

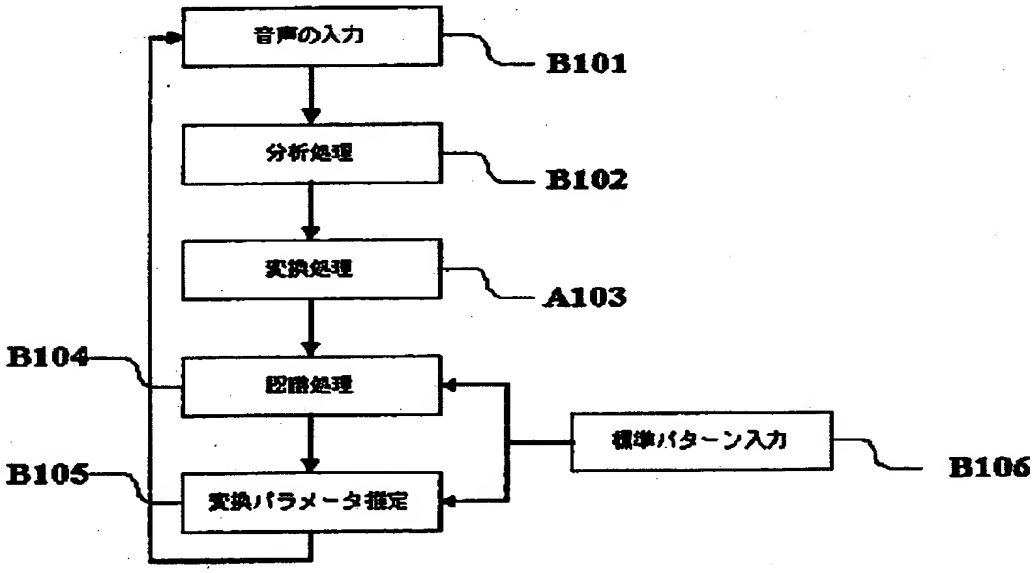
【図２】


音声の入力
A101
分析処理
A102
A103
変換パラメータの推定
標準パターンの入力
A105
変換処理
A104

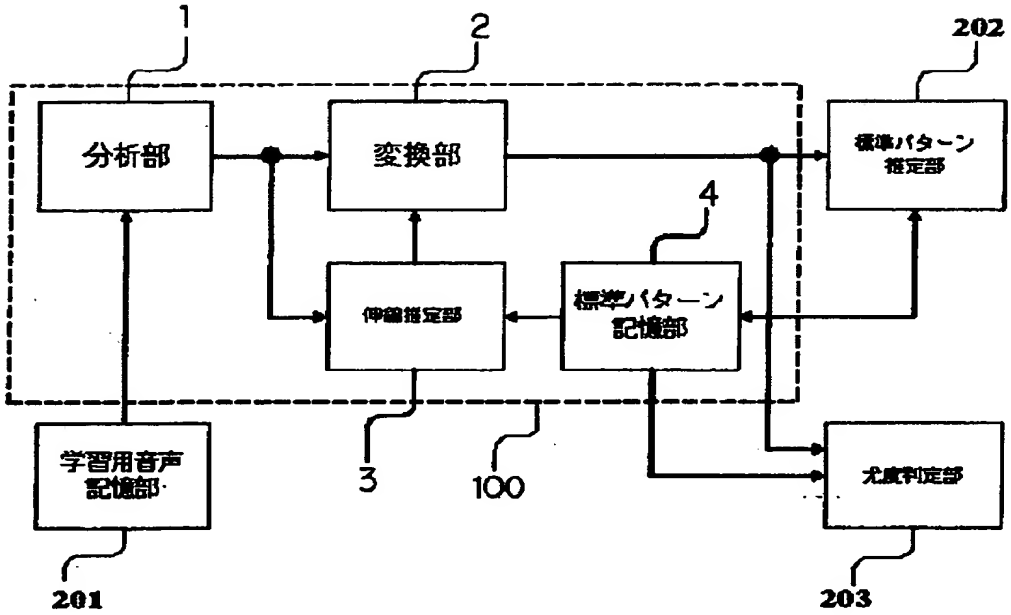
【図３】


1
2
101
分析部
変換部
認識部
伸縮推定部
標準パターン
記憶部
3
4
100

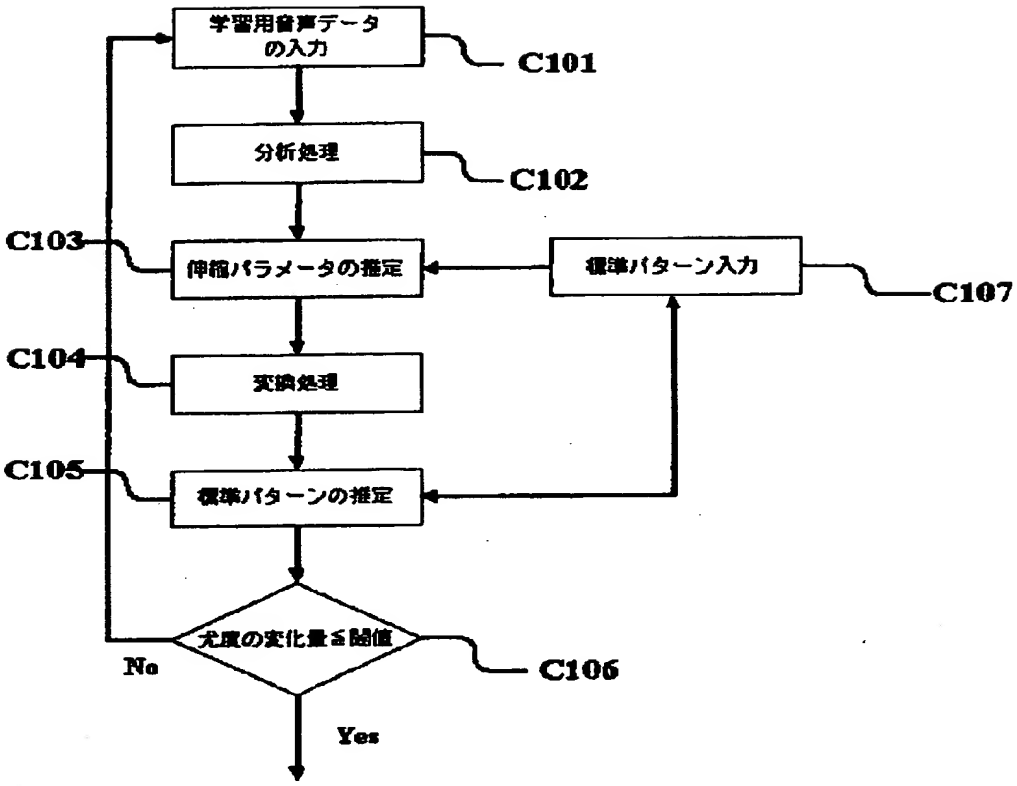
【図４】


音声の入力
B101
分析処理
B102
変換処理
A103
B104
認識処理
標準パターン入力
B106
B105
変換パラメータ推定

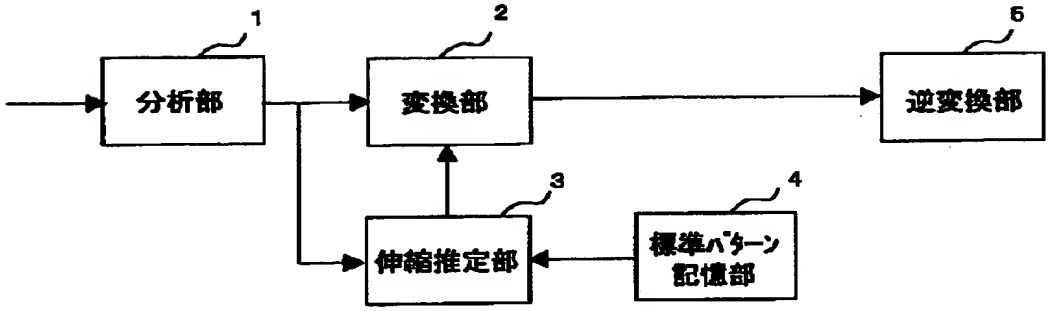
【図５】


1
2
202
分析部
変換部
標準パターン
推定部
4
伸縮推定部
標準パターン
記憶部
学習用音声
記憶部
3
100
尤度判定部
201
203

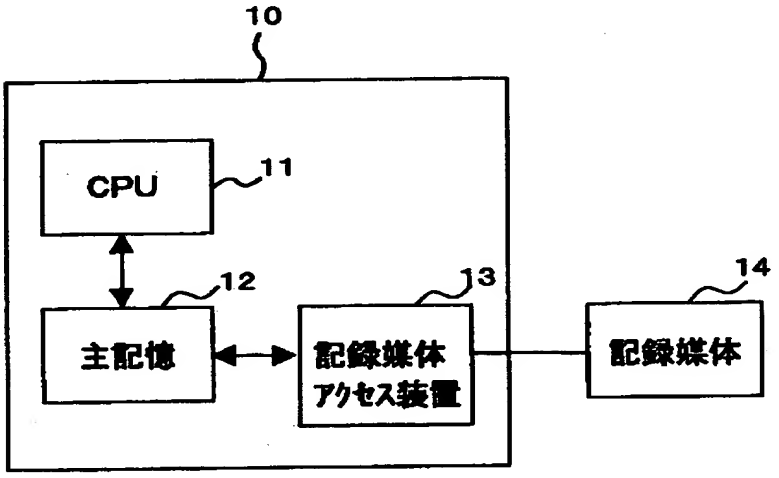
【図６】


学習用音声データの入力
C101
分析処理
C102
C103
伸縮パラメータの推定
標準パターン入力
C107
C104
変換処理
C105
標準パターンの推定
尤度の変化量≦閾値
C106
No
Yes

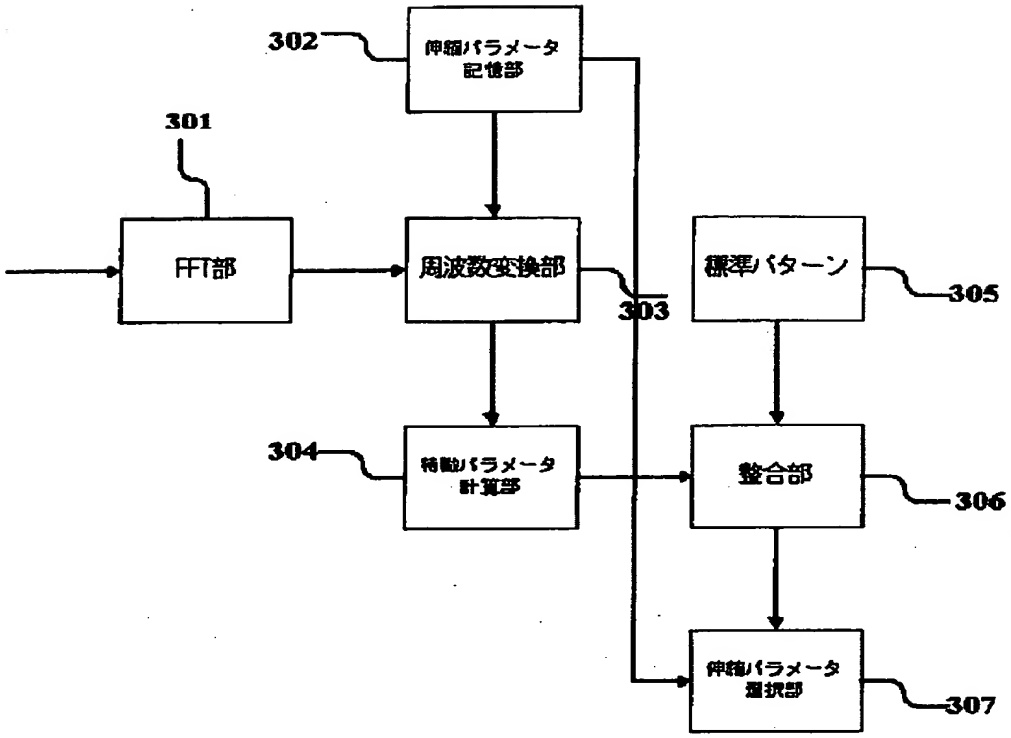
【図７】


1
分析部
2
変換部
5
逆変換部
3
伸縮推定部
4
標準ﾊﾟﾀｰﾝ
記憶部

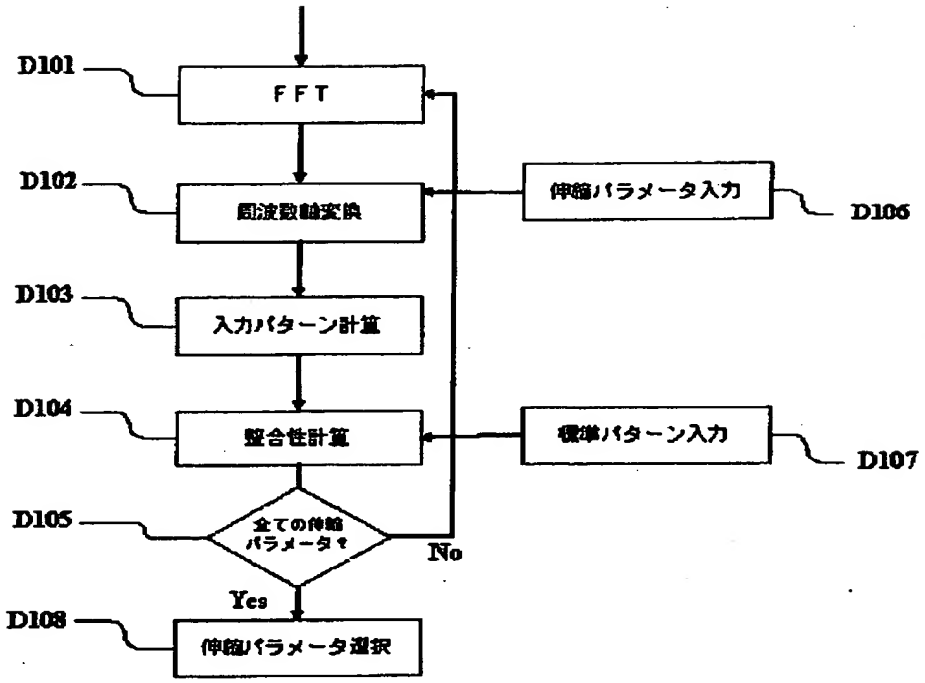
【図８】


10
CPU
11
12
主記憶
13
記録媒体
ｱｸｾｽ装置
14
記録媒体

【図９】


302
伸縮パラメータ記憶部
301
FFT部
周波数変換部
303
標準パターン
305
304
特徴パラメータ計算部
整合部
306
伸縮パラメータ選択部
307

【図１０】


D101
FFT
D102
周波数軸変換
伸縮パラメータ入力
D106
D103
入力パターン計算
D104
整合性計算
標準パターン入力
D107
D105
全ての伸縮パラメータ？
No
Yes
D108
伸縮パラメータ選択

【書類名】　　要約書

【要約】

【課題】

少ない演算量で話者毎に最適な伸縮パラメータの値を計算可能とし、性能を向上する装置及び方法の提供。

【解決手段】

入力音声信号をケプストラムを含む入力パターンに変換する分析部１と、標準パターンを記憶する標準パターン記憶部３と、入力パターンと標準パターンを用いて周波数軸方向の伸縮パラメータを出力する伸縮推定部４と、伸縮パラメータを用いて入力パターンを変換する変換部２と、変換部２によって変換された前記入力パターンと標準パターンとの距離を計算し最も距離の小さい標準パターンを認識結果として出力する認識部１０１を備え、伸縮推定部４は入力パターンに含まれるケプストラムを用いて伸縮パラメータを推定し、伸縮パラメータを決定する際に前もって様々な値を持つ必要が無く、様々な値について距離計算を行うことも不要とされる。

【選択図】

図３

出　願　人　履　歴　情　報

識別番号　　〔０００００４２３７〕

１．変更年月日　　１９９０年　８月２９日
［変更理由］　新規登録
住　所　東京都港区芝五丁目７番１号
氏　名　日本電気株式会社